IES Institute of Education Sciences

Effects of an Inquiry-Oriented Curriculum and Professional Development Program on Grade 7 Students' Understanding of Statistics and on Statistics Instruction

Regional Educational Laboratory Southeast

At Florida State University

REL 2021–055
U.S. DEPARTMENT OF EDUCATION

A Publication of the National Center for Education Evaluation and Regional Assistance at IES

Effects of an Inquiry-Oriented Curriculum and Professional Development Program on Grade 7 Students' Understanding of Statistics and on Statistics Instruction

Robert C. Schoen and Sharon Koon

January 2021

On average, Florida students earn only half of the points possible in the statistics content area of the state's annual mathematics assessment. Leaders in Broward County Public Schools, a large, diverse, urban school district, viewed changes to statistics curriculum and instruction as one way to address this issue. This study randomly assigned 40 middle schools in the district to either implement a replacement curriculum unit with four days of teacher professional development in probability and statistics or continue with their practice-as-usual instruction in probability and statistics. The replacement unit supported teaching and learning of all the probability and statistics standards in the grade 7 course description. The replacement unit with the associated professional development, called the Supporting Teacher Enactment of the Probability and Statistics Standards program, improved student understanding of statistics and statistics instruction. The magnitude of the effect on student understanding was 23 percent of 1 standard deviation, which is comparable to an increase of 9 percentile points for an average student.

Why this study?

Florida students struggle with the statistics content area of the state's annual mathematics assessment. In the 2018/19 school year—eight years after the state adopted the current mathematics curriculum standards—Florida students, on average, earned only half of the points possible in this area of the assessment—indicating that students are not mastering the content area as expected.

Concerns about low performance and inadequate implementation of the curriculum standards in statistics[1] led to conversations between the study team and mathematics curriculum leaders in Broward County Public Schools (BCPS). BCPS leaders hypothesized that grade 7 mathematics teachers might lack the requisite knowledge to teach statistics in a way that adequately prepares students to meet the state standards. These conversations led to the formation of a partnership comprising the study team, BCPS mathematics curriculum specialists, and representatives of the American Statistical Association (ASA).

The partnership identified a promising program for addressing the identified problems. The ASA publishes curriculum materials designed to support implementation of state probability and statistics standards. It also provides professional learning opportunities in statistics for secondary-level teachers in conjunction with its annual Meeting-Within-a-Meeting program. Members of the partnership agreed that a combination of these two components—the curriculum and associated professional development—held promise for improving the implementation of the statistics standards and, therefore, student learning in statistics. The resulting combination was named the Supporting Teacher Enactment of the Probability and Statistics Standards (STEPSS) program.

The lesson plans in the ASA curriculum materials use an instructional design model for teaching statistics that involves students in four phases of statistical investigation: formulating a question that can be answered by data, designing

For additional information, including background on the study and technical methods, access the report appendixes at https://go.usa.gov/xAXRK.

1. For brevity, "statistics" in this report refers to probability and statistics.

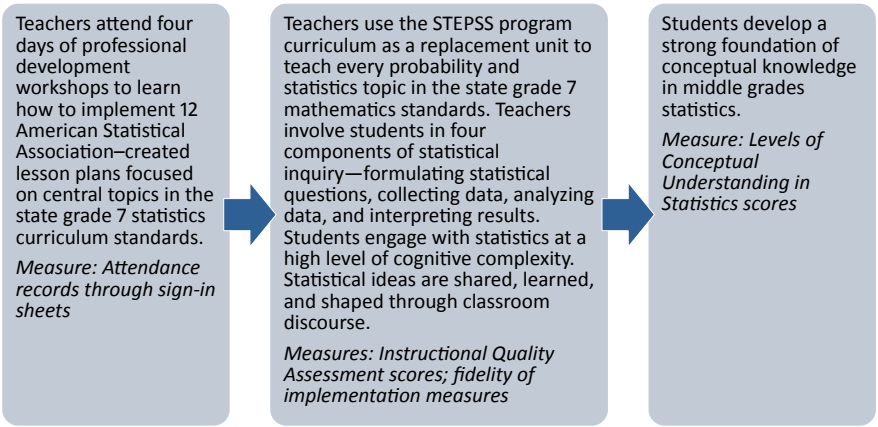
and implementing a data collection plan, analyzing the data using graphical and numerical methods, and interpreting the analyses in the context of the original question (Franklin et al., 2007). Each lesson plan is available in a book or web-based repository of curriculum resources curated by the ASA. The set of lesson plans make up the curriculum unit used in the STEPSS program, which starts with a brief review of key topics from grade 6 and addresses all the grade 7 probability and statistics curriculum standards. Although the lesson plans have existed for several years, their effect on student achievement or statistics instruction has not been evaluated (R. Nichols, personal communication, October 26, 2016).

This report describes a randomized controlled trial that evaluated the effect of the STEPSS program on student understanding of statistics and on statistics instruction. Schools assigned to the treatment condition used the program's 20-day curriculum unit, which addresses each standard in the probability and statistics domain for grade 7 mathematics in Florida, along with the associated teacher professional development workshops. Schools assigned to the comparison condition continued with practice-as-usual instruction and professional development. The study tested the STEPSS program in BCPS, a large, diverse, urban school district. (See appendix A for more information about the design and implementation of the study, including the components of the STEPSS program.)

The STEPSS theory of change describes the core components of the program that are expected to improve the teaching and learning of statistics (figure 1). The theory hypothesizes that these conceptually oriented, inquiry-based components help students develop a deeper understanding of fundamental ideas in statistics. In the STEPSS program, teachers attend two days of professional development workshops in the summer and two days during the school year. All four workshops occur before teachers are expected to teach the statistics unit in class. During the workshops teachers learn statistics content by participating in the same lessons that they will implement with students in their classrooms. Teachers play the role of students in the first three days of trainer-led workshops. In the last workshop teachers work together to prepare to teach the lessons to their own students. Small groups of teachers deliver the lessons to fellow teachers, who play the role of students. Teachers then reflect on the experience as a group and make final preparations to implement the lessons in their own classrooms. After the four workshops teachers implement the lessons as the statistics curriculum unit in their classrooms. Teachers can fit the unit into any portion of the school year.

Outcomes of interest in the study are student understanding of statistics and statistics instruction. The study team used the Levels of Conceptual Understanding in Statistics (LOCUS) test to measure student understanding of

Figure 1. Theory of change for the Supporting Teacher Enactment of the Probability and Statistics Standards program



Source: Authors' illustration.

statistics and the Instructional Quality Assessment (IQA; Boston, 2012) to measure statistics instruction (see box 1 for definitions of key terms). The LOCUS test measures student understanding of the foundational concepts in statistics and is aligned with the middle grade statistics standards in the Common Core State Standards (Jacobbe, 2016; Jacobbe et al., 2014).

The IQA is an observational measure of classroom instruction that measures the cognitive complexity of student tasks and the extent to which students are engaged in classroom discourse. Cognitive complexity of tasks ranges from recall of known facts or procedures at the low end of the scale to devising approaches to solve complex, novel problems or providing elaborate explanation, analysis, synthesis, or generalizations based on evidence at the high end. Low classroom discourse means no speaking by students or students providing brief, one-word answers to teachers' questions. High classroom discourse means that students engage in class discussions by providing evidence for claims, explaining their thinking, and comparing or contrasting ideas shared by different students.

This report discusses statistically significant results from these analyses and, where appropriate, provides additional descriptive details that support the findings. (More information about the IQA and how the study team used it is provided in appendix A.)

Box 1. Key terms

Baseline equivalence. The extent to which, at the start of the study, students in schools assigned to the Supporting Teacher Enactment of the Probability and Statistics Standards (STEPSS) program have characteristics similar to those of students in schools assigned to practice as usual.

Cluster-randomized controlled trial. A study in which groups (clusters) of individuals are randomly assigned to participate in a program to study the causal effect of the program on outcomes of interest. In education research cluster-randomized controlled trials may identify whole classrooms or schools as the clusters of individuals to be assigned. In this study the clusters comprised schools.

Differential attrition rate. The absolute value of the difference in the percentage of data that are missing from the randomized STEPSS program group and the randomized practice-as-usual group in the analytic sample used to estimate treatment effects.

Effect size. The magnitude of the difference in the average outcome between interventions as a proportion of the pooled standard deviation of students in STEPSS program schools and practice-as-usual schools—the two intervention conditions in this study. The effect size estimate used in this study is Hedges's g, following What Works Clearinghouse guidance (U.S. Department of Education, 2017).

Fidelity of implementation. The extent to which prespecified features of the planned intervention were implemented in a manner that was consistent with the design or plan. This can involve aspects such as adherence to the plan and quality of delivery. It may also include an analysis of the contrast in experiences of the people in the treatment and the comparison conditions.

Improvement index. The average difference in an outcome between interventions in terms of percentile rank (U.S. Department of Education, 2017). In this study the improvement index reflects the expected change in percentile rank of an average student in a practice-as-usual school if the student had been in a STEPSS program school.

Instructional Quality Assessment (IQA). An observational tool that measures several components of classroom instruction organized into two broad categories: cognitive complexity of student tasks and classroom discourse (Boston, 2012; Junker et al., 2005). The study team observed a single day of classroom probability or statistics instruction for each participating teacher in STEPSS program schools and practice-as-usual schools and used the IQA to score the cognitive complexity and classroom discourse.

Levels of Conceptual Understanding in Statistics (LOCUS) test. A test of both student and teacher understanding of statistics created by the LOCUS project with funding from the National Science Foundation to develop valid and reliable assessments

(Jacobbe et al., 2014). In this study participating students were administered the 23-item, paper-and-pencil, Beginner/Intermediate, Form 1 of the LOCUS test.

Overall attrition rate. The percentage of data that are missing in the analytic sample used to estimate treatment effects.

Percentile rank. The percentage of scores that fall at or below a given score on an outcome.

What Works Clearinghouse (WWC) standards. A widely accepted framework for rating studies in education research. This study used the guidance outlined in the WWC standards (U.S. Department of Education, 2017) to report on overall and differential attrition, student effect sizes, and improvement indices.

Research questions

This study analyzed the extent to which the STEPSS program affects student understanding of statistics and statistics instruction for grade 7 students in BCPS in Florida. Two research questions guided the analyses:

1. What is the effect of the STEPSS program on student understanding of statistics, as measured by the LOCUS test?

2. What is the effect of the STEPSS program on statistics instruction, as measured by the IQA?

The study's data sources, sample, and methods are summarized in box 2 and detailed in appendix B.

Box 2. Data sources, sample, and methods

Data sources. Broward County Public Schools (BCPS) provided demographic information and grade 6 scores on the Florida Standards Assessment Mathematics for students whose parents consented to sharing data with the study team after random assignment. After teaching the statistics unit, teachers BCPS administered the Levels of Conceptual Understanding in Statistics (LOCUS) test to each grade 7 mathematics student. The teachers then provided the study team with the responses from students with parental consent to participate in the study. The study team conducted classroom observations of statistics lessons in grade 7 classrooms in spring 2019 using the Instructional Quality Assessment (IQA). The study team selected one class for each teacher of regular or advanced grade 7 mathematics and observed the class on one day of probability or statistics instruction. LOCUS scale scores were the outcome measure used to answer research question 1 on the effect on student understanding of statistics, and IQA scores were the outcome measure used to answer research question 2 on the effect on statistics instruction. Classroom observers recorded the source of the enacted lesson or lessons and the phases of statistical investigation enacted during the lesson to measure fidelity of implementation. See appendix A for a detailed description of the outcome measures.

Sample. In May 2018 the study team, in partnership with BCPS, identified 40 regular middle schools eligible to participate in the study. To be eligible, schools had to have at least 30 grade 7 mathematics students and two or more grade 7 mathematics teachers (see appendix A for additional details on their identification). In spring 2018 the study team randomly assigned 20 schools to the Supporting Teacher Enactment of the Probability and Statistics Standards (STEPSS) program and 20 schools to practice as usual. The fall 2018 class rosters indicated the presence of 155 grade 7 mathematics teachers and 14,045 grade 7 mathematics students in the 40 schools.

The study lost a substantial proportion of schools, teachers, and students to attrition. The school district's Institutional Review Board required active parental consent for an individual student's demographic information and scores to be used in the study. Most parents did not return the consent forms. The analytic sample for research question 1 on the effect of the STEPSS program on student understanding of statistics includes 31 middle schools, 86 grade 7 mathematics teachers, and 2,283 grade 7 mathematics students. The overall attrition rate among schools for this research question is 23 percent (see table B1 in appendix B). School-level attrition differs by 15 percent between STEPSS program schools and practice-as-usual schools. In the 31 schools in the analytic sample the student nonresponse rate is 79 percent. Scores on the grade 6 Florida Standards Assessment Mathematics for

the analytic sample indicate baseline equivalence between students in STEPSS program schools and students in practice-as-usual schools (see table B2 in appendix B). This combination of overall attrition, differential attrition, and baseline equivalence of student achievement results in a study that has the potential to meet What Works Clearinghouse (WWC) group design standards with reservations (version 4.1) for the outcome of student understanding of statistics.

The analytic sample for research question 2 on the effect of the STEPSS program on statistics instruction includes 26 middle schools and 74 grade 7 teachers. The overall attrition rate among schools for this research question is 35 percent (see table B3 in appendix B). School-level attrition differs by 10 percent between STEPSS program schools and practice-as-usual schools. In the 26 study schools the teacher nonresponse rate was 30 percent. Data to examine equivalence of pretreatment IQA scores of the two groups of schools are not available. This combination of overall attrition, differential attrition, and absence of opportunity to investigate baseline equivalence results in a study that would likely not be deemed to meet WWC group design standards for the outcome of statistics instruction.

Methods. The study team used a cluster-randomized controlled trial to estimate the effect of the STEPSS program on student understanding of statistics and on statistics instruction (see appendix B for more information). The analytic approach (multilevel statistical analyses) accounts for the nesting of students within teachers and teachers within schools. To guard against overalignment of the outcome measure for student understanding of statistics (LOCUS scale scores) with the treatment (the STEPSS program), the study team observed the content of the professional development workshops and reviewed the curriculum materials and LOCUS test items to determine whether the test contained the same, or similar, tasks or situations as the workshops or curriculum. The items on the LOCUS test were judged to be independent from the tasks and situations in the STEPSS program. As an additional safeguard against bias, the student response data that test scorers received did not have any information about treatment condition.

Findings

This section discusses the findings of the study, starting with the average effect of the STEPSS program on student understanding of statistics (research question 1). It then reports the effect on statistics instruction (research question 2).

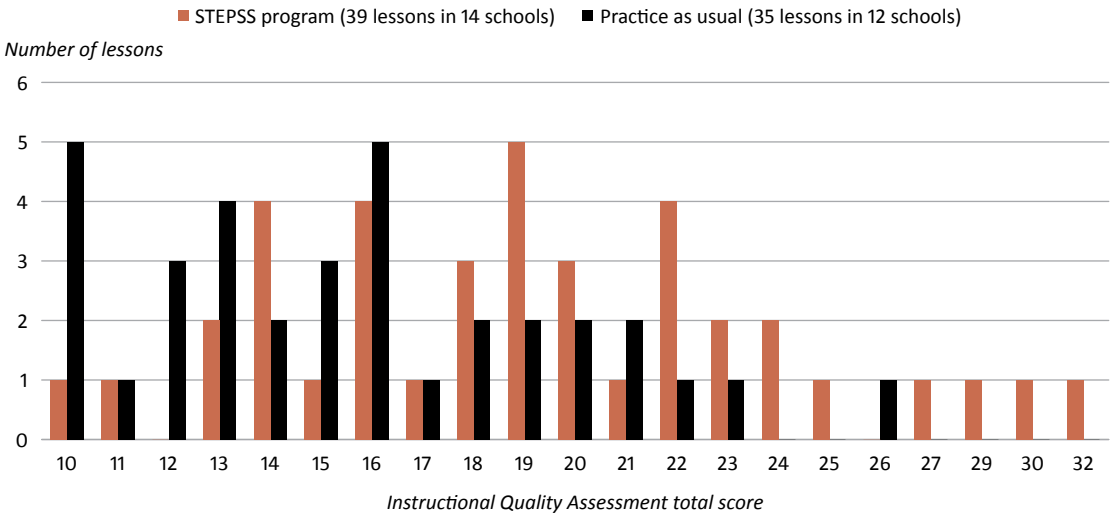
Students in schools assigned to the Supporting Teacher Enactment of the Probability and Statistics Standards program scored higher on the Levels of Conceptual Understanding in Statistics test than students in schools assigned to practice as usual

Grade 7 mathematics students in STEPSS program schools scored higher on the LOCUS test than their peers in practice-as-usual schools. The mean difference in scores is equal to an effect size of 0.23, which corresponds to an improvement index of +9 (see table B6 in appendix B). The magnitude of this change is equivalent to a student at the 50th percentile improving to the 59th percentile. This result indicates that the STEPSS program is likely to increase student understanding of statistics.

Statistics instruction in schools assigned to the Supporting Teacher Enactment of the Probability and Statistics Standards program received higher scores on the Instructional Quality Assessment than statistics instruction in schools assigned to practice as usual

On average, statistics instruction in STEPSS program schools involved students in tasks with higher cognitive complexity and engaged students in higher reasoning and discussion of each other's ideas about statistics, as measured by the IQA, than did statistics instruction in practice-as-usual schools (figure 2). The mean difference in scores is equivalent to an effect size of 0.80, which corresponds to an improvement index of +29 (see table B8 in appendix B).

Figure 2. Cognitive complexity and classroom discourse in statistics instruction were higher in schools assigned to the Supporting Teacher Enactment of the Probability and Statistics Standards program than in schools assigned to practice as usual, Broward County Public Schools, 2018/19



STEPSS is Supporting Teacher Enactment of the Probability and Statistics Standards.

Note: The overall median score on the Instructional Quality Assessment was 17; 64 percent of lessons in STEPSS program schools received scores above the median, compared with 34 percent of lessons in practice-as-usual schools.

Source: Authors' compilation of results from the study-administered Instructional Quality Assessment.

The lowest observed IQA total score was 10, and the highest was 32. Statistics lessons in practice-as-usual schools received scores in the lower half of the range (see figure 2). Low scores indicate that tasks assigned to students involved memorizing facts or practicing procedures prescribed by the teacher or instructional materials. Low scores also indicate that classroom discussion was absent, did not involve students or the teacher in revoicing or restating the statements made by teachers or students, or did not involve students in providing reasons or evidence to support claims that they made—if they made any claims at all. High scores indicate that students were engaged in tasks involving statistical inquiry, exploration of conjectures and generalizations, reasoning and explanations, or drawing connections between concepts and procedures or among different representations of the same idea. High scores also indicate that the teacher or students described how ideas or positions shared during the discussion relate to each other and explained their thinking or evidence for their claims.

The median total IQA score for the 74 lessons was 17. Almost two-thirds of lessons in STEPSS program schools received scores above the median, compared with one-third in practice-as-usual schools. The most separation occurred in the highest scores: 13 lessons in STEPSS program schools received scores above the third quartile, compared with 3 lessons in practice-as-usual schools.

Statistics instruction in schools assigned to the STEPSS program consistently involved the lessons provided in the program's curriculum unit. Teachers used the STEPSS program curriculum in 95 percent of the lessons observed in STEPSS program schools. Lessons from the STEPSS program curriculum were never observed in practice-as-usual schools, though teachers in practice-as-usual schools used sources other than the district-adopted mathematics curriculum in more than half of the observed lessons. This measure of fidelity of implementation indicates a sharp treatment contrast with respect to the instructional materials used to teach statistics by the schools assigned to the practice-as-usual or STEPSS program.

Each lesson plan in the STEPSS program curriculum is designed to involve all four phases of statistical investigation, as described in *Guidelines for Assessment and Instruction in Statistics Education* (Franklin et al., 2007).

Table 1. Number and percentage of observed statistics lessons that involved students in each of the four phases of statistical investigation, by condition, Broward County Public Schools, 2018/19

	STEPSS program		Practice as usual	
Phase of statistical investigation	Number	Percent	Number	Percent
Formulate a statistical question	13	33	2	6
Generate data	38	97	14	40
Analyze data	38	97	34	97
Interpret results	5	13	6	17

STEPSS is Supporting Teacher Enactment of the Probability and Statistics Standards.

Note: Observations occurred in 39 lessons in 14 STEPSS program schools and in 35 lessons in 12 practice-as-usual schools.

Source: Authors' analysis of data from classroom observations.

Classroom observers recorded whether the lesson enacted each phase in order to provide additional descriptive information about implementation of the inquiry-oriented instructional design model that is central to the theory of instruction in the STEPSS program. This provides another metric that can be used to examine the fidelity of implementation. Almost all statistics lessons in both STEPSS program schools and practice-as-usual schools involved students in the analyze data phase, which includes practices that mathematics teachers recognize as quantitative and mathematical in nature (table 1). Almost all lessons in STEPSS program schools involved students in the generate data phase, compared with 40 percent of lessons in practice-as-usual schools. The percentage of lessons that involved students in the formulate a statistical question phase was higher in STEPSS program schools than in practice-as-usual schools. Lessons in both STEPSS program schools and practice-as-usual schools infrequently involved students in the interpret results phase.

Implications

The study findings suggest that BCPS could consider implementing the STEPSS program districtwide to improve student understanding of and instructional practice in grade 7 statistics. The STEPSS program had a large, positive effect on student understanding of statistics—0.23 standard deviation. Mathematics programs that are subjected to randomized controlled trials rarely result in positive effects on student performance of that magnitude (Garet et al., 2016; Gersten et al., 2014; Yoon et al., 2007). In a summary of evaluations of 67 programs supported by the U.S. Department of Education's Investing in Innovation program the two largest effect size estimates were 0.20 and 0.23, and the median effect size was 0.03 (Boulay et al., 2018). Approximately half of the published results of randomized controlled trials in mathematics result in negative effects or effects that are not discernible from zero (Kraft, 2020). On average, students' overall mathematics achievement on nationally normed tests increases by about 0.23 standard deviation in grade 7 (Hill et al., 2008), suggesting that the STEPSS program may have doubled students' rate of learning of statistics compared with their peers in practice-as-usual schools. The findings place the STEPSS program on the short list of mathematics programs that have been subjected to a randomized controlled trial and found to have a positive effect on student achievement and classroom instruction.

The STEPSS curriculum unit is easily implemented in schools. It addresses all the benchmarks in the content standards in the regular and advanced grade 7 mathematics course descriptions. In Florida, content standards in the probability and statistics domain account for approximately one-third of the standards in the grade 7 mathematics course description, but the STEPSS curriculum unit requires only four weeks to implement—or 10–15 percent of the school year. It can be used to replace existing curriculum in the short term. In the longer term curriculum developers could consider incorporating the instructional design principles in the STEPSS program lessons into their own curricula for this domain.

Rough calculations based on the costs incurred in this study suggest a per student cost of $13.14 (see appendix A). Based on this estimate and the effect on student understanding of statistics, the STEPSS program is categorized as a low-cost, large-effect-size program (Kraft, 2020).

Limitations and future research

The study findings are subject to several limitations related to generalizability. The validity of the results is highest for the large, urban school district in which the study was conducted. Similar results may not be observed in other settings. Additional research might help determine the extent to which the findings would emerge in other contexts (for example, different grade levels, different school districts, or different accountability systems). The study focuses on the average effect for assenting students with parental consent to participate in the randomized controlled trial. Variation in effects across subgroups of schools, teachers, or students was not investigated. The effect size estimates for the STEPSS program focus solely on statistics, and the study cannot determine the extent to which the program influences student achievement or classroom instruction in other domains of the grade 7 mathematics curriculum. Further research is warranted on the replicability, generalizability, and possible variation in treatment effects of the findings.

This study focuses on relatively short-term outcomes because the LOCUS tests were administered shortly after the end of the statistics unit, and many of the statistics units were postponed for weeks or months from the recommended pacing guide. Changing the implementation of the STEPSS program to teach the statistics unit earlier in the school year would allow time for examination of other important outcomes, including performance on a delayed posttreatment test and the end-of-the-year state mathematics assessment. Future research is needed to examine other longer-term outcomes, including teachers' instructional practice after professional development support ends, student performance in statistics and in other domains of the mathematics or science curriculum in subsequent school years, coursetaking in high school statistics, and possibly even attitudes about statistics.

Additional research is needed on the various components of the multifaceted STEPSS program. Future research could explore the independent roles or contributions of the program's two interrelated components: curriculum resources and professional development for teachers. The ASA provides the lesson plans free of charge, but the professional development workshops are more expensive. It is unknown how teachers might have implemented the curriculum materials if they had not participated in the professional development. Further research is needed to determine whether the curriculum materials could yield similar results when teachers do not have a similar opportunity for professional learning. The extent to which more, less, or different support for implementation is required (or sufficient) and how teachers who require various supports can be identified are also important questions. A better understanding of the role of various facets of classroom instruction as mediators of the program's effect on student understanding of statistics might also yield important insights into the teaching and learning of statistics.

Several aspects of the study design might influence the magnitude of the effect size estimates. First, there is overlap among the personnel who developed the STEPSS curriculum, delivered the professional development workshops, and developed the LOCUS test. Although the study team ensured that there was not overalignment by reviewing whether items on the LOCUS test were used in the STEPSS program, the coherence could magnify the treatment effect. Second, the timing of the administration of the LOCUS test, which was administered shortly after the end of the curriculum unit and across a two-month testing window from April to May districtwide, could inflate effect size estimates. Finally, professional development programs with voluntarily participation tend to have larger effects than programs with mandatory participation (Kennedy, 2016), but the schools in the STEPSS program were informed by the district that they were participating.

School, teacher, and student attrition is a major issue for the study. High student nonresponse was driven in part by the fact that the district's Institutional Review Board required active, informed parental consent and student assent. Teachers in STEPSS program schools were more cooperative with requests to schedule classroom observations than teachers in practice-as-usual schools. The analysis of the effect on student understanding of statistics has the potential to meet WWC standards with reservations despite the high school-level attrition and high student nonresponse rates because of the similarity in pretreatment mathematics achievement between students in STEPSS program schools and students in practice-as-usual schools. The analysis of the effect on statistics instruction would likely not be deemed to meet WWC standards because there were no pretreatment data available on the outcome.

References

Boston, M. (2012). Assessing instructional quality in mathematics. *The Elementary School Journal, 113*(1), 76–104.

Boulay, B., Goodson, B., Olsen, R., McCormick, R., Darrow, C., Frye, M., et al. (2018). *The Investing in Innovation Fund: Summary of 67 evaluations. Final report* (NCEE No. 2018–4013). U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance.

Franklin, C., Kader, G., Mewborn, D., Moreno, J., Peck, R., Perry, M., et al. (2007). *Guidelines for assessment and instruction in statistics education report: A pre-K–12 curriculum framework.* American Statistical Association.

Garet, M. S., Heppen, J. B., Walters, K., Parkinson, J., Smith, T. M., Song, M., et al. (2016). *Focusing on mathematical knowledge: The impact of content-intensive teacher professional development* (NCEE No. 2016–4010). U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance. http://ies.ed.gov/ncee/pubs/20164010/pdf/20164010.pdf.

Gersten, R., Taylor, M. J., Keys, T. D., Rolfhus, E., & Newman-Gonchar, R. (2014). *Summary of research on the effectiveness of math professional development approaches* (REL 2014–010). U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Southeast. http://ies.ed.gov/ncee/edlabs.

Hill, C. J., Bloom, H. S., Black, A. R., & Lipsey, M. W. (2008). Empirical benchmarks for interpreting effect sizes in research. *Child Development Perspectives, 2*(3), 172–177.

Jacobbe, T. (2016). *Levels of Conceptual Understanding in Statistics (LOCUS) 2016 test summary report*. University of Florida.

Jacobbe, T., Case, C., Whitaker, D., & Foti, S. (2014). Establishing the content validity of the LOCUS assessments through evidence centered design. In K. Makar & R. Gould (Eds.), *Proceedings of the 9th International Conference on Teaching Statistics*. International Statistical Institute. Retrieved October 28, 2020, from https://icots.info/9/proceedings/home.html.

Junker, B., Matsumura, L. C., Crosson, A., Wolf, M. K., Levison, A., Weisberg, Y, et al. (2005, April). *Overview of the Instructional Quality Assessment*. Paper presented at the Annual Meeting of the American Educational Research Association.

Kennedy, M. M. (2016). How does professional development improve teaching? *Review of Educational Research, 86*(4), 945–980.

Kraft, M. A. (2020). Interpreting effect sizes of education interventions. *Educational Researcher, 49*(4), 241–253.

U.S. Department of Education, Institute of Education Sciences, What Works Clearinghouse. (2017). *What Works Clearinghouse Standards Handbook Version 4.0.*

Yoon, K. S., Duncan, T., Lee, S. W. Y., Scarloss, B., & Shapley, K. (2007). *Reviewing the evidence on how teacher professional development affects student achievement* (REL 2007–033). U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Southwest. http://eric.ed.gov/?id=ED498548.

REL 2021–055

January 2021

This report was prepared for the Institute of Education Sciences (IES) under Contract ED-IES-17-C-0011 by the Regional Educational Laboratory Southeast administered by the Florida Center for Reading Research, Florida State University. The content of the publication does not necessarily reflect the views or policies of IES or the U.S. Department of Education, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government.

Schoen, R. C., & Koon, S. (2021). *Effects of an inquiry-oriented curriculum and professional development program on grade 7 students' understanding of statistics and on statistics instruction* (REL 2021–055). U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Southeast. http://ies.ed.gov/ncee/edlabs.

This report is available on the Regional Educational Laboratory website at http://ies.ed.gov/ncee/edlabs.